ABSTRACT

A search engine for searching a corpus improves the relevancy of the results by classifying multiple terms in a search query as a single semantic unit. A semantic unit locator of the search engine generates a subset of documents that are generally relevant to the query based on the individual terms within the query. Combinations of search terms that define potential semantic units from the query are then evaluated against the subset of documents to determine which combinations of search terms should be classified as a semantic unit. The resultant semantic units are used to refine the results of the search.